

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 16:13:48 ; Search time 14 Seconds
(without alignments)
956,918 Million cell updates/sec

Title: US-09-988-292-16

Perfect score: 1759
Sequence: 1 MAYVAPAGYQPTYNPTLPY.....FORVDLEIQDVLTVYQI 323

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1759	100.0	323	1	LEG4_HUMAN
2	1419	80.7	323	1	LEG4_PIG
3	1367.5	77.7	324	1	LEG4_RAT
4	1227	69.8	301	1	LEG6_MOUSE
5	590	33.5	353	1	LEG9_MOUSE
6	588	33.4	354	1	LEG9_MOUSE
7	574	32.6	355	1	LEG9_MOUSE
8	531.5	30.2	316	1	LEG8_HUMAN
9	513.5	29.2	316	1	LEG8_MOUSE
10	507.5	28.9	316	1	LEG8_MOUSE
11	445.5	25.3	283	1	LEG1_HAECO
12	422	24.0	279	1	LEG3_CABEL
13	400	22.7	297	1	LEG3_CABEL
14	338.5	19.2	336	1	LEG7_HUMAN
15	283.5	16.1	135	1	LEG7_MOUSE
16	277.5	15.8	135	1	LEG7_MOUSE
17	275.5	15.7	144	1	LEG5_RAT
18	272.5	15.5	135	1	LEG7_HUMAN
19	272.5	15.5	244	1	LEG3_CRILO
20	264.5	15.0	261	1	LEG3_RAT
21	263.5	15.0	261	1	LEG3_MOUSE
22	255	14.5	295	1	LEG3_MOUSE
23	241.5	13.7	249	1	LEG3_HUMAN
24	240	13.6	241	1	LEG3_RABIT
25	174.5	9.9	134	1	LEG6_CHICK
26	172.5	9.8	134	1	LEG4_CHICK
27	171.5	9.7	129	1	LEG4_HUMAN
28	162	9.2	129	1	LEG4_MOUSE
29	157	8.9	180	1	LEG8_MOUSE
30	156	8.9	180	1	LEG8_MOUSE
31	153.5	8.7	130	1	LEG2_MOUSE
32	153	8.7	134	1	LEG1_SHEEP
33	153	8.7	1216	1	YPM5_CABEL

34	152	8.6	179	1	LEG7_CABEL
35	151	8.6	134	1	LEG1_MOUSE
36	150	8.5	134	1	LEG1_BOVIN
37	147	8.4	134	1	LEG1_CRIGR
38	146	8.3	134	1	LEG1_RAT
39	140.5	8.0	130	1	LEG2_RAT
40	133.5	7.6	123	1	LEG2_PIG
41	132.5	7.5	132	1	LEG2_HUMAN
42	127.5	7.2	134	1	LEG1_BUPAR
43	116	6.6	135	1	LEG1_CONMY
44	102.5	5.8	135	1	LEG2_CONMY
45	97.5	5.5	105	1	LEGX_MOUSE

ALIGNMENTS

RESULT 1
LEG4_HUMAN STANDARD; PRT; 323 AA.
ID LEG4_HUMAN
AC P56470;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galectin-4 (lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36LBP) (Antigen NY-CO-27).
GN LGALS4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric carcinoma;
RA Kato S.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA MEDLINE=97454307; PubMed=9310382;
RT Rechreche H., Mallo G.V., Montalto G., Dagorn J.C., Iovanna J.L.;
RT Cloning and expression of the mRNA of human galectin-4, an S-type
RT lectin down-regulated in colorectal cancer.;
RT Eur. J. Biochem. 248:225-230(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=97306339; PubMed=9162064;
RT Huffelt M.B., Jordan E.T., Gilt M.A., Barondes S.H., Leflier H.;
RT "Strikingly different localization of galectin-3 and galectin-4 in
RT human colon adenocarcinoma T84 cells. Galectin-4 is localized at sites
RT of cell adhesion.";
RT J. Biol. Chem. 272:14294-14303(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Skin;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; AB006781; BAA22165.1; -

```

DR EMBL; AF014838; AAC51763.1; -.
DR EMBL; U82953; AAB86590.1; -.
DR EMBL; BC003661; AAH03661.1; -.
DR EMBL; BC005146; AAH05146.1; -.
DR HSP; P47929; 1BKZ.
DR Genew; HGNC:6565; LGALS4.
DR MIM; 602518; -.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat.
FT DOMAIN 1 152 GALAPTIN 1.
FT DOMAIN 153 177 LINKER.
FT DOMAIN 178 323 GALAPTIN 2.
FT BINDING 256 262 BETA-GALACTOSIDE (BY SIMILARITY).
FT BINDING 256 262 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 323 AA; 35941 MW; 579BC0A9AB3990EF CRC64;

Query Match 100.0%; Score 1759; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.6e-137;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPAPGYQPTNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
DB 1 MAYVPAPGYQPTNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

QY 61 AFHNPFRDGDWVKVFNLTLOGGKWSSEERKSRMPFKGAFAFELFVFLAHHYKVVNGNP 120
DB 61 AFHNPFRDGDWVKVFNLTLOGGKWSSEERKSRMPFKGAFAFELFVFLAHHYKVVNGNP 120

QY 121 FYEGHRLPLQWTHLVQDGLQSLQINFIQGOPLRPGQPPMPYPYPGHCHQOOLNSLP 180
DB 121 FYEGHRLPLQWTHLVQDGLQSLQINFIQGOPLRPGQPPMPYPYPGHCHQOOLNSLP 180

QY 121 FYEGHRLPLQWTHLVQDGLQSLQINFIQGOPLRPGQPPMPYPYPGHCHQOOLNSLP 180
DB 121 FYEGHRLPLQWTHLVQDGLQSLQINFIQGOPLRPGQPPMPYPYPGHCHQOOLNSLP 180

QY 181 TMGEPPTFNPPVYFGRLOGLTARRTIIIGKVPPTGKSFAINFKVSSGDIALHINPR 240
DB 181 TMGEPPTFNPPVYFGRLOGLTARRTIIIGKVPPTGKSFAINFKVSSGDIALHINPR 240

QY 241 MGNGTVRNSLLGWSGSEBKTIHNPFGQPFDSLIRCGLDRFKVYANGQHLFDFAH 300
DB 241 MGNGTVRNSLLGWSGSEBKTIHNPFGQPFDSLIRCGLDRFKVYANGQHLFDFAH 300

QY 301 LSAFORVDTLEIQGDVTLISVQI 323
DB 301 LSAFORVDTLEIQGDVTLISVQI 323

RESULT 2
LEG4 PIG STANDARD; PRT; 323 AA.
AC Q29058; Q29296.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36LBP).
GN LGALS4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=tongue;
RX MEDLINE=95081129; PubMed=7989350;
RA Chiu M.L., Parry D.A.D., Feldman S.R., Klapper D.G., O'Keefe E.J.;
RT "An adherens junction protein is a member of the family of lactose-
binding lectins."
RL J. Biol. Chem. 269:31770-31776 (1994).
[2]
RN
RP SEQUENCE OF 1-140 FROM N.A.
RC TISSUE=Small intestine;
RX MEDLINE=96327607; PubMed=8672129;

```

```

RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA
RL Mamm. Genome 7:509-517(1996).
CC -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.
CC -!- SUBUNIT: MONOMER.
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X79303; CAA55884.1; -.
CC EMBL; F14653; CAA23179.1; -.
CC HSP; P47929; 1BKZ.
CC InterPro; IPR001079; Galectin.
CC Pfam; PF00337; Gal-bind lectin; 2.
CC SMART; SM00276; GLECT; 2.
CC PROSITE; PS00309; GALAPTIN; FALSE_NEG.
KW Galaptin; Lectin; Repeat.
FT DOMAIN 1 152 GALAPTIN 1.
FT DOMAIN 153 177 LINKER.
FT DOMAIN 178 323 GALAPTIN 2.
FT BINDING 256 262 BETA-GALACTOSIDE (BY SIMILARITY).
FT CONFLICT 98 99 AP -> GA (IN REF. 2).
FT CONFLICT 126 126 H -> T (IN REF. 2).
SQ SEQUENCE 323 AA; 35852 MW; 728C761712B29343 CRC64;

Query Match 80.7%; Score 1419; DB 1; Length 323;
Best Local Similarity 80.5%; Pred. No. 2.1e-109;
Matches 260; Conservative 25; Mismatches 38; Indels 0; Gaps 0;

QY 1 MAYVPAPGYQPTNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
DB 1 MAYVPAPGYQPTNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

QY 61 AFHNPFRDGDWVKVFNLTLOGGKWSSEERKSRMPFKGAFAFELFVFLAHHYKVVNGNP 120
DB 61 AFHNPFRDGDWVKVFNLTLOGGKWSSEERKSRMPFKGAFAFELFVFLAHHYKVVNGNP 120

QY 121 FYEGHRLPLQWTHLVQDGLQSLQINFIQGOPLRPGQPPMPYPYPGHCHQOOLNSLP 180
DB 121 FYEGHRLPLQWTHLVQDGLQSLQINFIQGOPLRPGQPPMPYPYPGHCHQOOLNSLP 180

QY 181 TMGEPPTFNPPVYFGRLOGLTARRTIIIGKVPPTGKSFAINFKVSSGDIALHINPR 240
DB 181 TMGEPPTFNPPVYFGRLOGLTARRTIIIGKVPPTGKSFAINFKVSSGDIALHINPR 240

QY 241 MGNGTVRNSLLGWSGSEBKTIHNPFGQPFDSLIRCGLDRFKVYANGQHLFDFAH 300
DB 241 MGNGTVRNSLLGWSGSEBKTIHNPFGQPFDSLIRCGLDRFKVYANGQHLFDFAH 300

QY 301 LSAFORVDTLEIQGDVTLISVQI 323
DB 301 LSAFORVDTLEIQGDVTLISVQI 323

RESULT 3
LEG4 RAT STANDARD; PRT; 324 AA.
AC P38552;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36LBP).

```

```

OS LGALIS4.
ON Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=93194902; PubMed=844956;
RA Oda Y., Herrmann J., Gilt M., Turick C.W., Burlingame A.L.,
RA Baerondes S.H., Loeffler H.;
RT "Soluble lactose-binding lectin from rat intestine with two different
RL carbohydrate-binding domains in the same peptide chain.";
RT J. Biol. Chem. 268:5929-5939(1993).
CC CC - FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS.
CC CC - SUBUNIT: MONOMER.
CC CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FULL-LENGTH FORM IN SMALL
CC AND LARGE INTESTINE AND STOMACH BUT WAS NOT DETECTED IN OTHER
CC TISSUES INCLUDING LUNG, LIVER, KIDNEY, AND SPLEEN.
CC CC - DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC CC - SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, WJ353; AAA1505.1; -.
DR PIR, A46631; A46631.
DR HSSP, PA7929; IBKZ.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
DR Galaptin; Lectin; Repeat.
FT FT DOMAIN 1 152 GALAPTIN 1.
FT DOMAIN 153 177 LINKER.
FT DOMAIN 178 324 GALAPTIN 2.
FT BINDING 257 263 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 324 AA; 36347 MW; 478024D7322AFEB7B CRC64;

Query Match 77.7%; Score 1367.5; DB 1; Length 324;
Best Local Similarity 76.6%; Pred. No. 3.6e-105;
Matches 249; Conservative 33; Mismatches 40; Indels 3; Gaps 2

QY 1 MAVYPAGVGYPTNPTLPYYOPIPGGLANYSMSVTIYGVAASHMKRFVNVPVGDDPSDV 60
Db 1 MAVYPAGVGYPTNPTLPYKRPIPGSLSVGMSIYIGAIKDMNRFRHNFPAVGDEADI 60
QY 61 AFHNPFDFCDMDKVNTLTGGCKWSSEKKSMPPKGAFLAEFYIVLAETKYVVNGNP 120
Db 61 AFHNPFDFCDMDKVNTMTQSGMGKEEKSKMSPFOKHFEELVFVMWSEHYKVVNVGTP 120
QY 121 FVEYGHNLPLQMTHLOVDGDLQLOSINFGOLPLPCGPMM--PPVPGHGCHQOLNS 178
Db 121 FVEYGHNLPLQMTHLQVDDDELQSINFLLGQPAPASQTPGTMITIPALPASGVNPPOMS 180
QY 179 LPMTGEPTNPVPVYFGRLOGGLTARRRIIIIKGVVPTPKGSFAINFKVSGSSGIATAHN 238
Db 181 LPVWAGPPIPNPVVYVGTLIQGSLTARRRIIIIKGVVLPTANKLLINFKVSGTGDIATHNM 240
QY 239 PRMGNGTVVANSLSNGWSSEKKITHNPFQCGOFFDISIRCGLDREFKVVANGOHFDFA 298
Db 241 PRIGD-CVYNNSYMNGSWGSEERKIPLYNPFAGOGFFDISIRCGTRFKVFPANGOHLDFDS 299
QY 299 HRISAFORVDTEIOGDVTLTSYVOI 323
Db 300 HRFAFORVDMLETIGDITLTSYVOI 324

```

ID	LEG6 MOUSE	STANDARD	PRT	301 AA.
AC	054851	088352		
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Galectin-6.			
GN	LGAL6.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/Sv.			
RX	MEDLINE=98112847; PubMed=9446608;			
RA	Gilt M.A., Colnot C., Peitler F., Nani K.J., Baronides S.H.,			
RA	Leffler H.,			
RT	"galectin-4 and galectin-6 are two closely related lectins expressed			
RT	in mouse gastrointestinal tract.";			
RL	J. Biol. Chem. 273:2954-2960(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/Sv.			
RX	MEDLINE=98112848; PubMed=9446609;			
RA	Gilt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Baronides S.H.,			
RA	Leffler H.,			
RT	"sequence, structure, and chromosomal mapping of the mouse Lgal6			
RT	gene, encoding galectin-6.";			
RL	J. Biol. Chem. 273:2961-2970(1998).			
CC	- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING			
CC	DOMAINS.			
CC	- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF026796; AAC04508.1; -			
DR	EMBL; AF026796; AAC04508.1; JOINED.			
DR	EMBL; AF026797; AAC04508.1; JOINED.			
DR	EMBL; AF026798; AAC04508.1; JOINED.			
DR	EMBL; AF026794; AAC27244.1; -			
DR	HSSP; P17931; 1A3K.			
DR	MGD; MG1:107535; Lgal6.			
DR	InterPro; IPR001079; Galectin.			
DR	Pfam; PF00337; Gal-bind_lectin. 2.			
DR	SMART; SM00276; GLECY. 2.			
DR	PROSITE; PS00309; GALAPTIN; 2.			
DR	Galaptin, Lectin; Repeat.			
FT	DOMAIN 1 151 GALAPTIN 1.			
FT	DOMAIN 152 160 LINDER.			
FT	DOMAIN 160 301 GALAPTIN 2.			
FT	CONFICT 154 154 A -> V (IN REF. 2; AAC27244).			
SO	SEQUENCE 301 AA; 34112 MW; 9A4DD09944EDFAB9 CRC64;			
Query Match	69.8%; Score 1227; DB 1; Length 301;			
Best Local Similarity	70.3%; Pred. No. 1,je-93;			
Matches 227; Conservative 32; Mismatches 42; Indels 22; Gaps 3.				
Qy	1 MAYVPAAGVGYOPTNPTLPYQYIPGGLVGVMSVYIQGVASEHMKRFVNVVGVGDGSDV 60			
Db	1 MAYVPAAGVGYOPTNPTLPYKRIPEGGLVGVMSVFIQGVAKENMRFRVNVVAVGVDDGADV 60			
Oy	61 AFHFNPRFDGDKVYVNTLQGGKMGSEKRSKMPFKGAALFVLYLAHYVYVYVNGNP 120			

Db 61 AFHFNPRFGDQKVVFNKQSGRWGKEEK-SMPFGKRGHFLVFMVPEHYKVVVNGSP 119
 Qy 121 FYEGHRLPLQWTHLOVGDLOLQSNFTIGQPLRQPGPPMPPYPGPGHCHQQLNSLP 180
 Db 120 FYEGHRLPLQWTHLOVGDLOLQSNFTIGQPLRQPGPPMPPYPGPGHCHQQLNSLP 159
 Qy 181 TMGPGPTFPNPPVYFGRLOGLTARTIIKGYVPTGKSFAINFKVSGSDIALHINPR 240
 Db 160 AMTGPPVPNCLPVGALQGFVTRTIIKGYVPTGKSFAINFKVSGSDIALHINPR 219
 Qy 241 MNGTVVRNLLNGSGSEBKKTTHNPPGPGQFDDLSIRCLDRPKVYANGHLPFAHR 300
 Db 220 IGD-CLVRNSYMGSGWETRMVAYNPPFGQFDDLSIRCLDRPKVYANGHLPFAHR 278
 Qy 301 LSAFQVRDTLEIOGVDVTLNVQI 323
 Db 279 FQALRKINTLEINGDLTSLVHI 301

RESULT 5

LEG9 MOUSE STANDARD; PRT; 353 AA.

AC O08573; O08572;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Galectin-9.
 GN LGALS9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Small intestine, and Kidney;
 RX MEDLINE=97190351; PubMed=9038233;
 RA Wada J., Kanwar Y.S.;
 RT "Identification and characterization of galectin-9, a novel beta-galactoside-binding mammalian lectin."
 RL J. Biol. Chem. 272:6078-6086(1997).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=97298141; PubMed=9153289;
 RA Wada J., Ota K., Kumar A., Wallner E.I., Kanwar Y.S.;
 RT "Developmental regulation, expression, and apoptotic potential of galectin-9, a beta-galactoside binding lectin."
 RL J. Clin. Invest. 99:2452-2461(1997).
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-CLASSICAL SECRETORY PATHWAY.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: ACCENTUATED EXPRESSION IN LIVER AND THYMUS OF EMBRYO, DETECTED IN EMBRYONIC HEART, BRAIN, LIVER, AND KIDNEY, HIGHLY EXPRESSED IN ADULT THYMUS, SMALL INTESTINE, AND LIVER, AND TO A LESSER EXTENT IN LUNG, KIDNEY, SPLEEN, CARDIAC, AND SKELETAL MUSCLE. BARELY DETECTABLE IN BRAIN AND RETICULOCYTE.
 CC THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
 CC -!- DEVELOPMENTAL STAGE: THE EXPRESSION INCREASED WITH SUCCESSIVE STAGES OF EMBRYONIC DEVELOPMENT.
 CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U55061; AAB51190.1; -.
 DR EMBL; U55060; AAB51189.1; -.
 DR HSP; P17931; IAK3.
 DR MGD; MGI:109496; Igals9.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal-bind_lectin; 2.
 DR SMART; SM00276; GLECT; 2.
 DR PROSITE; PS00309; GALAPTIN; 2.
 KW Galaptin; Lectin; Repeat; Alternative splicing.
 FT DOMAIN 1 147 GALAPTIN 1.
 FT DOMAIN 148 204 LINKER.
 FT DOMAIN 205 353 GALAPTIN 2.
 FT BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).
 FT BINDING 285 291 BETA-GALACTOSIDE (BY SIMILARITY).
 FT VARSPLIC 148 178 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 353 AA; 40036 MW; B54036FE280C531 CRC64;
 Query Match 33.5%; Score 590; DB 1; Length 353;
 Best Local Similarity 38.7%; Pred. No. 2.7e-41;
 Matches 135; Conservative 46; Mismatches 126; Indels 42; Gaps 9;
 Qy 11 PTYNPTLPYQPIPGGLNVGMSVYIQVASEHMKRFFVNFVVGQDPSDVAFHFNPRFDG 70
 Db 9 PYINPIPTGPIQGLQGLQVLTQTTKSFARFVNF-QNSFNGNDIAFHFNPRFEE 67
 Qy 71 WDKVVFNTLOGGKWSSEERKSMPEKKAAGAFELVFLVLAHYKVVVNGNPFYEGHRLPL 130
 Db 68 GGYVVCNTKQNGQWGPBERKQMPFQKMPFELCLVQRSEKFMVNVKFFVQYQHRVPY 127
 Qy 131 QMVTHLQVDDGLQIQSINFIG-----GQPLR-----PQPPMPPYPGPGH- 171
 Db 128 HLVTVIAVSGCLKLSFITFQNSAAPVQHVFTLQFSQVQFPRTPKGRKQKTONFRPAHQ 187
 Qy 172 -----CHQ---QLNSLPTWEGPPTNP-----PVYFGRLOGLTARTIIKGYV 214
 Db 188 APMAQTTHMVHSTPGQMFSTPGI--PPVYVTPAVTIPFTPIPNGLYPSKSIMISGNV 245
 Qy 215 PPTCKSFALNPKVSGSDIALHINPRMNGCTVVRNLLNGSGSEKKIT-HNPPGPGQF 273
 Db 246 LPDTRRHINLRG--GDIAFHLNPRFNENAVVRNTQINNNGWGEERSLGRMPFSGRQS 303
 Qy 274 FDLISIRGLDRFKVYANGHLPFAHLSAFQVRDTLEIOGVDVTLNVQI 322
 Db 304 FSVWICEGHCFKVAVNGQHCVEYHKLKQLDINTLEAVAGDIQLTHVQ 352
 RESULT 6
 ID LEG9 RAT STANDARD; PRT; 354 AA.
 AC P97840; O08588; O35866;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Galectin-9 (36 kDa beta-galactoside binding lectin) (Urate transporter/channel) (UAT).
 GN LGALS9.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney, and Small intestine;
 RX MEDLINE=97190351; PubMed=9038233;
 RA Wada J., Kanwar Y.S.;
 RT "Identification and characterization of galectin-9, a novel beta-galactoside-binding mammalian lectin."
 RL J. Biol. Chem. 272:6078-6086(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney;
 RX MEDLINE=97150769; PubMed=8995305;
 RA Leal-Pinto E., Tao W., Rappaport J., Richardson M., Knorr B.A.,

```

RA Abramsos R.G.;
RT "Molecular cloning and functional reconstruction of a urate
RT transporter/channel."
RT J. Biol. Chem. 272:617-625(1997).
CC -1- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYOCYTE-
CC BETHHELMAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY
CC SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS
CC CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE
CC METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC
CC TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND
CC GASTROINTESTINAL URATE EXCRETION. HIGHLY SELECTIVE TO THE ANION
CC URATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-
CC CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE
CC SMALL INTESTINE.
CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U59462; AAB51192.1; -.
CC EMBL; U72741; AAB68592.1; -.
CC EMBL; U67958; AAB48591.1; -.
CC DR HSBP; P17931; IAKK.
CC DR InterPro; IPR001079; Galactin.
CC Pfam; PF00337; Gal-bind lectin; 2.
CC SMART; SMO0276; GLECT; 2.
CC DR PROSITE; PS00309; GALAPTIN; 2.
CC KW Galaptin; Lectin; Repeat; Alternative splicing; Ion transport.
CC FT DOMAIN 1 147 GALAPTIN 1.
CC FT DOMAIN 148 205 LINKER.
CC FT DOMAIN 206 354 GALAPTIN 2.
CC FT BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).
CC FT BINDING 286 292 BETA-GALACTOSIDE (BY SIMILARITY).
CC FT VASAPLIC 148 179 MISSING (IN SHORT ISOFORM).
CC SQ SEQUENCE 354 AA; 39946 MW; 6574F60B2EAF37C CRC64;
CC
CC Query March 33.4%; Score 588; DB 1; Length 354;
CC Best local similarity 38.8%; Pred. No. 4e-41;
CC Matches 138; Conservative 43; Mismatches 123; Indels 52; Gaps 10;
CC
CC QY 10 QPTV-NPTLPYQPIPIGGLNVMGYIIOGVSEHNKRRFVNFVGVGDGDSVAFFHFNPRF 68
CC Db 7 QPYNVNPVPIFGIILIOGGLQNGLOITLIGQTVHPFNRLAVNVQTFPS-GNDIAFFHFNPRF 65
CC
CC QY 69 DGMKVYVNTLLOGGKWSGEERKSNRPFKGAFAELVELYLAHYVYVNVGNDFYIYGRIL 128
CC Db 66 BEGGYVVCNTKNGKMGGEERKQMPFOKMPFELCFIVQSEPFVWVKQNFVQYSHRV 125
CC
CC QY 129 PLQWVTHLQVGGDLQLOGINFIG-----GQFLRPGG--P 160
CC Db 126 PYHLVDITISVSGCHLSIFINFNQNSTAARVQVFTMQSPQVQVPPRMKGRKQRTGGFOP 185
CC
CC QY 161 PMMPD-----YPGDGHCHQQLNLSLPTMGPEPTFNP-----PVYFGRLQGGTLART 207
CC Db 186 ALQAAVACTQIIHTVHSIP-----QMLSTGTGCI--PMAYATPAYTIPFTFISINGEYPSKS 239
CC
CC QY 208 IIKKIVYPTPGSGPAINFKVYSSGDIALHINPRMGNTYVNSLNGWSGESEKITT-HN 266
CC Db 240 INISGVVLPDARFPHINIRCG--GDIAFELINRFEKVKVVRNTQIINNSGPEERSIPGRM 297
CC
CC QY 267 PGQGFQPLSRGGLDRFKYVANGQHLFDFAHRLSAFQVDTLEIQGVLTLSYQ 322
CC

```

```

298 PFNNGQSF5WILCEGCFKVAVDGQHICEYVHRLKNLPDINTLEVDAGDIQJTHVQ 353
RESULT 7
ID LEGS HUMAN STANDARD; PRT; 355 AA.
AC 000182; 014512; 075028; G9N058;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galectin-9 (HOM-HD-21) (Ecalectin).
GN LGALS9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Spleen;
RX MEDLINE=67197815; PubMed=9045665;
RA Tureci O., Schmitt H., Fadle N., Freundschuh M., Sahin U.;
RT "Molecular definition of a novel human galectin which is immunogenic
in patients with Hodgkin's disease.";
RJ J. Biol. Chem. 272:6416-6422(1997).
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Gastric carcinoma;
RX Kato S.;
RN Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RX MEDLINE=98307937; PubMed=9642261;
RA Matsunoto R., Matsunoto H., Seki M., Hata M., Asano Y., Kanegasaki S.,
RA Stevens R.L., Hirashima M.;
RT "Human ecalectin, a variant of human galectin-9, is a novel eosinophil
chemoattractant produced by T lymphocytes.";
RJ J. Biol. Chem. 273:16976-16984(1998).
RN [4]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RA Akiyama S.;
RT "Homo sapiens galectin-9 (LGALS9) / ecalectin gene, exon 2 through
11.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RA Gressler J., Spitzberger F., Schroeder H.E.;
RT "Genomic organization of the human galectin-9 gene.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE
SHORT ISOFORM ACTS AS AN EOSINOPHIL CHEMOATTRACTANT.
RN [7]
RP ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
RN [8]
RP TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC
TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.
RN [9]
RP DOMAIN: CONTRAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
DOMAINS.
RN [10]
RP SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
RN [11]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch.
RN [12]
RP EMBL; Z49107; CAA89322.1; -;
RN EMBL; AB006782; BAA2156.1; -;
RN EMBL; AB005894; BAA31542.1; -;
RN EMBL; AB040130; BAB83625.1; -;
RN EMBL; AB040129; BAB83625.1; JOINED.
RN EMBL; AB040130; BAB83624.1; -;

```

```

DR EMBL; AB040129; BAB83624.1; JOINED.
DR EMBL; AJ288083; CAB93851.1; -.
DR EMBL; AJ288084; CAB93851.1; JOINED.
DR EMBL; AJ288085; CAB93851.1; JOINED.
DR EMBL; AJ288086; CAB93851.1; JOINED.
DR EMBL; AJ288087; CAB93851.1; JOINED.
DR EMBL; AJ288088; CAB93851.1; JOINED.
DR EMBL; AJ288089; CAB93851.1; JOINED.
DR EMBL; AJ288090; CAB93851.1; JOINED.
DR HSP; F17931; IAK3.
DR Genew; HGNC:6570; LGALS9.
DR MIM; 601879; -.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat; Alternative splicing.
DR DOMAIN 1 148 GALAPTIN 1.
FT DOMAIN 149 206 LINKER.
FT DOMAIN 207 355 GALAPTIN 2.
FT BINDING 82 88 BETA-GALACTOSIDE (BY SIMILARITY).
FT BINDING 287 293 BETA-GALACTOSIDE (BY SIMILARITY).
FT VARSPLIC 149 180 MISSING (IN SHORT ISOFORM).
FT CONFLICT 5 5 G -> S (IN REF. 3).
FT CONFLICT 48 48 N -> D (IN REF. 5).
FT CONFLICT 79 81 NGS -> KGR (IN REF. 5).
FT CONFLICT 88 88 K -> R (IN REF. 1).
FT CONFLICT 89 89 T -> M (IN REF. 5).
FT CONFLICT 135 135 S -> F (IN REF. 1).
FT CONFLICT 270 270 P -> L (IN REF. 1).
FT CONFLICT 313 313 E -> G (IN REF. 1).
FT CONFLICT 326 326 L -> V (IN REF. 5).
FT CONFLICT 341 341 R -> K (IN REF. 5).
SQ SEQUENCE 355 AA; 39518 MW; 4748C22FCAFA536A CRC64;

Query Match 32.6%; Score 574; DB 1; Length 355;
Best Local Similarity 38.2%; Pred. No. 5.7e-40;
Matches 134; Conservative 52; Mismatches 121; Indels 44; Gaps 10;

QY 11 PTYNPTLPYQPIGGNLNMGSVYIQG-VASEHMKRFFNVFVVGQDPSVAHFHNPRED 69
Db 9 PYSFPAVFFSGTIQGGLDQGLQITVNTGLSSGTRFAVNFQTGFS-GNDIAFHNPREF 67

QY 70 GWDKVVNTLQGGKSGSEERKSRMPKGAFLVFTVLAELHVKVVGNGVNPFEYGHRLP 129
Db 68 DGGVVCNTQNGSGWGEERKTEMPFGKMPFDLCFLVQSSDFKVMVNGLLFVQYFHRVP 127

QY 130 LQVYTHLVQVGDGLQLOLSINFIGGQ-----PLRPQG-----PPMPP 165
Db 128 FHRVDTISVNGSVQLSVISYFQNPRTVPVQPAFSTVPSPQVPCFPFRGRQKPPGVWPA 187

QY 166 YPGP-----GHCHQ-----QLNSLPTMEGPTFP-----PVYFGRLOGGLTARTIIKG 212
Db 188 NPAPITQVHTVQSPAGQFMFSTPAI--PPMYFHPAYPMPFTITLGLGLPSKSLLSG 245

QY 213 YVPRTGKSFAINFKVSGSGDIALHINPMGNGTVVRNLLNGSGSEKKITHN-PFGPG 271
Db 246 TVLPSAORFHIN--LCSGNHIAFLNPRFDENAVVRNTQIDNSWGSSEERSLPRKMPFVVG 303

QY 272 QFFDLSCRLDRFKVYANGHLFDFAHRLSFAFQVDTLEIQGDVLTYSVQ 322
Db 304 QSFQSVMLCEAHCLKVADQGHLYFYHRLNRLNPTLNRLNLEVGGDIQLTHVQ 354

RESULT 8
LEG8 HUMAN
ID LEG8 HUMAN STANDARD; PRT; 316 AA.
AC O0214; O15215; Q9UP34; Q9UEZ6; Q9UP33; Q9UP32; Q9H584; Q9H585;
AC Q96B92;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galectin-8 (Gal-8) (Prostate carcinoma tumor antigen 1) (PCTA-1)

```

```

DE (Po66 carbohydrate-binding protein) (Po66-CBP).
GN LGALS8
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=96293510; PubMed=8692978;
RA Su Z.-Z., Lin J., Shen R., Fisher P.E., Goldstein N.I., Fisher P.B.;
RT "Surface-epitope masking and expression cloning identifies the human
RT prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin
RT gene family.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:7252-7257(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=20438187; PubMed=10980616;
RA Hadari Y.R., Eisenstein M., Zakut R., Zick Y.;
RT "Galectin-8: on the road from structure to function.";
RN Trends Glycosci. Glycotechnol. 9:103-112(1997).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Lung carcinoma;
RA Brichory F., Bidon N., Desrues B., Bourguet P., Le Pennec J.P.,
RA Dazord L.;
RT "Molecular cloning of a beta-galactoside-binding lectin related to
RT galectin-8 and identified in human lung carcinoma.";
RN Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Maier C., Haeussler J., Roesch K., Moschgath E., Vogel W.;
RT "Genomic organization and expression of the human galectin-8 gene.";
RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=20438187; PubMed=10980616;
RA Gopalakrishnan R.V., Roberts T., Tuli S., Kang D., Christiansen K.A.,
RA Fisher P.B.;
RT "Molecular characterization of prostate carcinoma tumor antigen-1,
RT PCTA-1, a human galectin-8 related gene.";
RN Oncogene 19:4405-4416(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Lahm H., Siebert H.-C., Andre S., Hoeflich A., Diehl D., Sordat B.,
RA Galtner H., Wolf E., Gabius H.-J.;
RT "Coca (Colorectal carcinoma-derived) galectin-8 variant I full-length
RT cDNA from a human colorectal carcinoma cell line.";
RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RN Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain, and Skin;
RA Strausberg R.;
RN Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1/1 (shown here), 2 and 3; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: UNIKITOUS. SELECTIVE EXPRESSION BY PROSTATE
CC CARCINOMAS VERSUS NORMAL PROSTATE AND BENIGN PROSTATIC
CC HYPERTROPHY.
CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL, L78132; AAB51605.1; ALT_INIT.
DR EMBL, X91790; CAA62904.1; ALT_INIT.
DR EMBL, AF074000; MAD45402.1; -
DR EMBL, AF074001; MAD45403.1; -
DR EMBL, AF074002; MAD45404.1; -
DR EMBL, AF193805; AAF19370.1; ALT_INIT.
DR EMBL, AF193805; AAF19370.1; JOINED.
DR EMBL, AF342815; AAK16733.1; ALT_INIT.
DR EMBL, AL136105; CAC15946.1; ALT_INIT.
DR EMBL, AL136105; CAC15947.1; ALT_INIT.
DR EMBL, BC015818; AAI15818.1; -
DR EMBL, BC015818; AAI15818.1; -
DR HSSP, P17931; IAK3.
DR Genew, HGNC:6569; LGALS8.
DR MIM, 606099; -
DR InterPro, IPR001079; Gallectin.
DR Pfam, PF00337; Gal-bind lectin; 2.
DR SMART, SM00276; GLECT; 2.
DR PROSITE, PS00309; GALAPTIN; 1.
DR Galactin; Lectin; Repeat; Antigen; Alternative splicing; Polymorphism.
KW GALAPTIN 1.
FT DOMAIN 154 184
FT BINDING 185 316
FT BINDING 248 254
FT VARSPIC 182 182
FT VARSPIC 182 182
FT VARSPIC 182 182
FT VARIANT 18 18
FT VARIANT 35 35
FT VARIANT 55 55
FT VARIANT 55 55
FT CONFLICT 97 99
FT CONFLICT 111 111
FT CONFLICT 170 170
FT CONFLICT 183 183
FT CONFLICT 203 203
SQ SEQUENCE 316 AA; 35538 MW; 1724C36085A0510A CRC64;
Query Match 30.2%; Score 531.5; DB 1; Length 316;
Best Local Similarity 37.6%; Pred. No. 1.5e-36;
Matches 120; Conservative 42; Mismatches 132; Indels 25; Gaps 4;

QY 13 YNPFLPYOP1PGGLANGMSVYIGVASEHMKRPFVNVVGOD--PGSDVAFHFNPRFDG 70
DB 12 YNPFLPYOP1PGGLANGMSVYIGVASEHMKRPFVNVVGOD--PGSDVAFHFNPRFDG 71
QY 71 WDKVVENTLOGKMGSEERKSMPEFKGAAPFELYLAHYKVVVNGNPPREYGHRLPL 130
DB 72 AGCIVCNTLLNEKMGREBEITYDTPFKREKSEFELYIMVLKDKFQVAVANGKHTLLYGRHIC 131
QY 131 QMVTLLQVDDGLQSLNFIIGGQPLRPGQPPMPRPYRPGHCHQOQNSLP---TWEGRP 186
DB 132 EKIDTLLGIGKVNINHSIGFSSFDLQ-----STQASLLLEISRENPV 175
QY 187 TFNP--VFPGGLTGLTARTIILKGVVPTGKSFAINFKVSSGDIALHINPRMGN 243
DB 176 KSGFPLSLPFAARLINTPMGKRTVVVKGAVNANAKSFNVDLAAGKSDIALHINPRMGN 235
QY 244 GTVVRNSLNGWSGSEKKITTHNPPGQGFPLDSIRCGLDKRFKYVANGQHLFDFARLSA 303
DB 236 KAFVRNSFLQSGWGEERNTSPFSSGMYFEMITICDVRERKVAIVANGVSLLEKRFKE 295
QY 304 FQVDTLEIQGDVTLTVYQ 322
DB 296 LSSIDTLEINGDHLEVR 314

RESULT 9
LEG8 MOUSE STANDARD; PRT; 316 AA.
ID LEG8 MOUSE STANDARD; PRT; 316 AA.
AC 09JL15;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galectin-8 (LGALS-8).
GN LGALS8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA Maier C., Haussler J., Roesch K., Moschagath E., Haussler J.,
RA Vogel W.,
RT "The human Lgals-8 gene: genomic sequence and expression of the
RT prostate carcinoma tumour antigen (PCTA-1) and the P066 carbohydrate
RT binding protein."
RU Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: POSSIBLE SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL, AF218069; AAF27645.1; -
DR HSSP, P17931; IAK3.
DR MGD, MGI:1928481; Lgals8.
DR InterPro, IPR001079; Galectin.
DR Pfam, PF00337; Gal-bind lectin; 2.
DR SMART, SM00276; GLECT; 2.
DR PROSITE, PS00309; GALAPTIN; 1.
KW Galactin; Lectin; Repeat.
FT DOMAIN 1 153
FT DOMAIN 154 184
FT DOMAIN 185 316
FT BINDING 248 254
SQ SEQUENCE 316 AA; 36161 MW; 11A20309AEF52C69 CRC64;
Query Match 29.2%; Score 513.5; DB 1; Length 316;
Best Local Similarity 36.0%; Pred. No. 4.6e-35;
Matches 111; Conservative 49; Mismatches 137; Indels 11; Gaps 3;

QY 13 YNPFLPYOP1PGGLANGMSVYIGVASEHMKRPFVNVVGOD--PGSDVAFHFNPRFDG 70
DB 12 YNPFLPYOP1PGGLANGMSVYIGVASEHMKRPFVNVVGOD--PGSDVAFHFNPRFDG 71
QY 71 WDKVVENTLOGKMGSEERKSMPEFKGAAPFELYLAHYKVVVNGNPPREYGHRLPL 130
DB 72 AGCIVCNTLLNEKMGREBEITYDTPFKREKSEFELYIMVLKDKFQVAVANGKHTLLYGRHIC 131
QY 131 QMVTLLQVDDGLQSLNFIIGGQPLRPGQPPMPRPYRPGHCHQOQNSLP---TWEGRP 190
DB 132 EKIDTLLGIGKVNINHSIGFSSFDLQ-----STQASLLLEISRENPV 175
QY 187 TFNP--VFPGGLTGLTARTIILKGVVPTGKSFAINFKVSSGDIALHINPRMGN 243
DB 176 KSGFPLSLPFAARLINTPMGKRTVVVKGAVNANAKSFNVDLAAGKSDIALHINPRMGN 235
QY 244 GTVVRNSLNGWSGSEKKITTHNPPGQGFPLDSIRCGLDKRFKYVANGQHLFDFARLSA 310
DB 236 KAFVRNSFLQSGWGEERNTSPFSSGMYFEMITICDVRERKVAIVANGVSLLEKRFKE 295
QY 304 FQVDTLEIQGDVTLTVYQ 322
DB 296 LSSIDTLEINGDHLEVR 314


```

Db    303   SVDGDRL 310
      :||:|
RESULT 10
LEG8 RAT
ID LEG8_RAT STANDARD; PRT; 316 AA.
AC Q62665;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Galectin-8 (30 kDa S-type lectin) (RL-30).
GN LGALS8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=95155445; PubMed=7852431;
RA Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y.;
RT "Galectin-8. A new rat lectin, related to galectin-4.";
RJ J. Biol. Chem. 270:3447-3453(1995).
CC -1- FUNCTION: POSSESSES SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, CARDIAC MUSCLE,
CC LUNG, AND BRAIN.
CC -1- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN WHOLE EMBRYOS, HIGH LEVELS
CC IN ADULT TISSUES.
CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC ENBL; U09824; AAA66359.1; --
CC HSSP; P17931; IA3K.
CC InterPro; IPR001079; Galectin.
CC Pfam; PF00337; Gal-bind lectin; 2.
CC SMART; SM00276; GLEGT; 2.
CC PROSITE; PS00309; GALAPTIN; 1.
KW Galaplin; Lectin; Repeat.
FT DOMAIN 1 153 GALAPTIN 1.
FT DOMAIN 154 184 LINKER.
FT DOMAIN 185 316 GALAPTIN 2.
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 36038 MW; C04B766CFE913D59 CRC64;
Query Match 28.9%; Score 507.5; DB 1; Length 316;
Best Local Similarity 34.7%; Pred. No. 1.4e-34;
Matches 111; Conservative 45; Mismatches 129; Indels 35; Gaps 3;
QY 13 YNPTLPYYQIPGGLNVGMVSIVTCGVASEHKMRFFVFNFVGQD--PGSDVAFHNPFRFDG 70
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Dd 12 YNPTIPVSVITTEQLKGSLIVIRGHVPKDSERFQVDFOHGNSLKPRAADVAFHNPFRKR 71
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 71 WDKVFNFTLGQGWKSSEERKSRMPFKGAAPFLVFIVLAEHYKVVNNGPNFYEGHRLPL 130
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Dd 72 SNCIVCNLTINEKGWGEIEITHDMPEFRKEGFVIWLVKNKFHVAVNGKIILYAHRNP 131
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 131 QMVTHLQVGDGLQLSINFIGGOLPRQPQPMPPYPGPGCHCQOOLSFLT----- 181
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Dd 132 EKIDTLGIFGKVINHSIGF-----RFSSDLQSMETSTLTGLTQIS 170
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 182 ---MEGGPTNPVPYFGRLOGGLTARRTHIIIKGYVPPTGKSFAINPKVSGSDIALHIIN 238
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



```

Db 153 PVPYSGIAGFPIDKTLIFGVYKAKKRFNINL-LRNGDIALHFNPFDEKAVIRNA 211
QY 251 LUNGSGWSEKIKTHNPFQGFPLSRGCLDRPKVYANGHLPFAHLSAFQVDTL 310
Db 212 LAANWENGERE-GKNPFKGVGFPLAKNEAYAFQIFVNGEFTSFARHODP-NDISGL 269
QY 311 EIQGDVTLSTYVOI 323
Db 270 QIQGDIELTGIOI 282

RESULT 12
LE32 CAEEL STANDARD; PRT; 279 AA.
ID LE32 CAEEL
AC P36573;
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 32 kDa beta-galactoside-binding lectin (32 kDa GBP).
GN W09H1.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348399; PubMed=1639789;
RA Hirabayashi J., Satoh M., Kasai K.-I.;
RT "Evidence that Caenorhabditis elegans 32-kDa beta-galactoside-binding
RT protein is homologous to vertebrate beta-galactoside-binding lectins.
RT cDNA cloning and deduced amino acid sequence.";
RL J. Biol. Chem. 267:15485-15490(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476274; PubMed=9334250;
RA Arata Y., Hirabayashi J., Kasai K.-I.;
RT "Structure of the 32-kDa galectin gene of the nematode Caenorhabditis
RT elegans";
RL J. Biol. Chem. 272:26669-26677(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Snye R.;
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 202-220.
RX MEDLINE=92348337; PubMed=1639749;
RA Hirabayashi J., Satoh M., Ohyama Y., Kasai K.-I.;
RT "Purification and characterization of beta-galactoside-binding
RT proteins from Caenorhabditis elegans.";
RL J. Biochem. 111:553-555(1992).
CC -1- FUNCTION: BINDS GALACTOSE.
CC -1- PFM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M94671; AAB87718.1; -
DR EMBL; AB000802; BAA22942.1; -
DR EMBL; Z82081; CAB04959.1; -
DR PIR; S27798; S27798.
DR PIR; PX0062; PX0062.
DR PIR; A42846; A42846.
DR HSSP; P17931; 1A3K.
DR WormPep; W09H1.6a; CE16576.
DR InterPro; IPR001079; Galectin.

```

```

DR Pfam; PF00337; Gal_bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat; Multigene family.
FT DOMAIN 1 146 GALAPTIN 1.
FT DOMAIN 2 147 279 GALAPTIN 2.
FT BINDING 213 219 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 279 AA; 31809 MW; ED9AE2AE3757IDA CRC64;

Query Match 24.0%; Score 422; DB 1; Length 279;
Best Local Similarity 33.0%; Pred. No. 1,3e-27;
Matches 101; Conservative 53; Mismatches 118; Indels 44; Gaps 9;

QY 10 OPTNPTLPYQPIPGINVGSVYIQCVASEHMKRFVNF--VVGDDPGSDVAFHNPR 67
Db 5 EPKSYF-VPYRSVLQEKREPGQTLIVKSTIDESGRFTINLSKADPSGNDVPLHVSVR 63
QY 68 FQGWKRVFNTIQGKMSSEKRSMPFKGAPELVIVLAHKKVYVNNPNPYEGHR 127
Db 64 FDE-GKIVLNSFSNGEWEKERK-SNPIKKDSDPDIRIRAHDDRQIIVDKERFDEHR 121
QY 128 LPLQWTHLQVVDLQLOSINIFIGGQPLRPGPPMPPYPPGCHQQLNSLPTMEGPT 187
Db 122 LPLSLSHLSITGDLXLYNHVH-GSKYV----- 148
QY 188 FNPVPYFGRLOGGLTARTIIKGYVPTGKSFAINFKVSSGDIALHINPRMGNTV 247
Db 149 ---PVPYSGIANGLPVGSLLVFGTVKAKARFHVNL-LRNGDISPHFNPRDEKAVI 204
QY 248 RNSLUNGSGWSEKIKTHNPFQGFPLSRGCLDRPKVYANGHLPFAHLSAFQV 307
Db 205 RNSLANWENGERE-GKNPFKGVGFPLVQNEBYAFQVFNGBERTYSFAHRAHPHD-I 262
QY 308 DTLEIQGDVTLSTYVOI 323
Db 263 AGLQISGDIELSGIOI 278

RESULT 13
LE33 CAEEL STANDARD; PRT; 297 AA.
ID LE33 CAEEL
AC 009581; O9GNP6;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 32 kDa beta-galactoside-binding lectin lec-3 (32 kDa GBP).
GN LEC-3 OR ZK892.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Hirabayashi J., Hayama K., Kasai K.;
RT "Novel galectins found in C. elegans.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Lloyd C.;
RT Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Jones S.J.M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS GALACTOSE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

```
CC -----
CC EMBL; AB038504; BAB11969.1; -.
CC EMBL; Z48638; CA88570.2; -.
CC HSSP; P47929; 1BKZ.
CC WormPep; ZK892.1; CE24743.
CC InterPro; IPR001079; Galectin.
CC Pfam; PF00337; Gal-bind lectin; 2.
CC SMART; SM00276; GLECT; 2.
CC PROSITE; PS00309; GALAPTIN; 1.
KW Galaptin; Lectin; Repeat; Multigene family.
FT DOMAIN 1 137 GALAPTIN 1.
FT DOMAIN 138 297 GALAPTIN 2.
FT BINDING 224 230 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 237 AA; 33641 MW; A48PCED3D8644082 CRC64;

Query Match 22.7%; Score 400; DB 1; Length 297;
Best Local Similarity 31.5%; Pred. No. 8.8e-26;
Matches 102; Conservative 52; Mismatches 110; Indels 60; Gaps 11;

QY 17 LPYQPIPGGLNMGSVYIQGVASEHMKRFFVNFVVGOD----PGSDVAHFHPRF 72
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 9 IPYRSKUTEREPQOTLIIRKTDIESKRFNIN--LHKDSPDFSGNDVPLHLSTIRFDE-G 65
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 73 KVFENTLOGGKMGSEERKSMFPGKGAFAELFVTLAEHYKVVVNGNPFVEYGHRLPQM 132
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 56 KIVNAYTKGTWGKEERAKN-PIKGGDDFDIRIRAHDSKQFVSNHKEVKNFHEHRLPNS 124
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 133 VTHLQVGDGLQLOLSINFTGGQPLRPOGPMMPYPGPGCHCHQLNSLPTMEGPPTEPNPV 192
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 125 VSHLSIDGVDVLNVHQM-GGKYV-----PV 148

QY 193 PY-FGRLOGGLTARTIIIKGYVPTGKSFAINFKVSSGDIHAIHINPRMN----- 243
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 149 PYESGIAADGLVPGKTLVYVTEPKAKFNINL-LKNGDIALHFNPRDEKANGFMCA 207
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 244 ---GTVVRNLLNGSWGSEKKITHTNPFPGQPFDFLIRCGLDLPKPKVYANGHQLDFDAH 299
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 208 KPTGGSVVRNLLNGSWGSEKKITHTNPFPGQPFDFLIRCGLDLPKPKVYANGHQLDFDAH 299
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 300 RLSAFQVDTLEIQGVDTLSVYQI 323
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 267 RVDPHD-IAGLIQIGDIETLGIQV 289
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 14
LEGC HUMAN STANDARD; PRT; 336 AA.
ID Q96DT0; Q96DS9; Q9NZ02; Q96PR9; Q9H259; Q9H258;
AC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galectin-12 (Galectin-related inhibitor of proliferation).
OS Homo sapiens (Human).
GN LGALS12 OR GRIP1.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D).
RC TISSUE=Adipose tissue;
RX MEDLINE=21423969; PubMed=11435439;
RA Hotta K., Funahashi T., Matsukawa Y., Takahashi M., Nishizawa H.,
RA Kishida K., Matsuda M., Kuriyama H., Kihara S., Nakamura T.,
RA Tochino Y., Bodkin N.I., Hansen B.C., Matsuzawa Y.;
RT "Galectin-12, an adipose-expressed galectin-like molecule possessing
RT apoptosis-inducing activity.";
RL J. Biol. Chem. 276:34089-34097(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS B; E AND F).
RC TISSUE=Retina;
RX MEDLINE=21283005; PubMed=11283015;
```

```
RA Yang R.-Y., Hsu D.K., Yu L., Ni J., Liu F.-T.;
RT "Cell cycle regulation by galectin-12, a new member of the galectin
RT superfamily.";
RL J. Biol. Chem. 276:20252-20260(2001).
CC -!- FUNCTION: Binds lactose.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 6 isoforms; a (shown here), b/GRIP1a
CC c, d, e/1, and f/2; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Not widely expressed. Predominantly expressed
CC in adipose tissue.
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF244974; AAK77328.1; -.
CC EMBL; AF244975; AAK77329.1; -.
CC EMBL; AF244976; AAK77330.1; -.
CC EMBL; AF244977; AAK77331.1; -.
CC EMBL; AF310686; AAG40863.1; -.
CC EMBL; AF310687; AAG40864.1; -.
CC EMBL; AF222695; AAF34677.1; -.
CC Genew; HGNC:15788; LGALS12.
CC MIM; 606096; -.
CC InterPro; IPR001079; Galectin.
CC Pfam; PF00337; Gal-bind lectin; 1.
CC SMART; SM00276; GLECT; 1.
CC PROSITE; PS00309; GALAPTIN; FALSE NEG.
KW Galaptin; Lectin; Repeat; Nuclear protein; Alternative splicing.
FT DOMAIN 47 183 GALAPTIN 1.
FT DOMAIN 194 336 GALAPTIN 2.
FT VARSPLIC 1 22 MISSING (IN ISOFORM E AND ISOFORM F).
FT VARSPLIC 1 61 MISSING (IN ISOFORM B AND ISOFORM D).
FT VARSPLIC 74 74 H -> HS (IN ISOFORM F).
FT VARSPLIC 200 208 MISSING (IN ISOFORM C AND ISOFORM D).
SQ SEQUENCE 336 AA; 37541 MW; F35D0A0CE509E795 CRC64;

Query Match 19.2%; Score 338.5; DB 1; Length 336;
Best Local Similarity 30.7%; Pred. No. 1.1e-20;
Matches 98; Conservative 46; Mismatches 136; Indels 39; Gaps 8;

QY 11 PTYNPTLPYQPIPGGLNMGSVYIQGVASEHMKRFFVNFVVGOD--PGSDVAHFHPRF 68
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 41 PVFHPVPVVTTFGGLHAGQWVLMQGVVPLDAHRFQVDFQCGLCLPRPDIAFHPRF 100
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 69 DGW-DKVVNTLOGGKMGSEERKSMFPGKGAFAELFVTLAEHYKVVVNGNPFVEYGH 127
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 101 HTTKPHVICNTLHGGRQREARWPHLALRGSSPLILFLFGNEVKVYVNGQHLHFRYR 160
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 128 LPLQWTHLQVDDGLQLOLSINFTGGQPL-----RPGPPMMPYPGPGCHCHQLNSLPT 181
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 161 LPLSHVDTLGIFGDIIVEAGFLNINPFVEGSEYPAAGHPF-----LLMSPR 207
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 182 MEGPPTFNPPVPYFGRLOGGLTAERTIIIKGYVPTGKSFAINFKVSSGDIHAIHINPRM 241
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 208 LE-----VPCSHALPQGLSPQGVIIIVRGLVLOEPKGFHTVSLR-----DQAAHAPVTL 254
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 242 GNGTVVRNLLNGSWGSEKKITHTNPFPGQPFDFLIRCGLDLPKPKVYANGHQLDFDAH 299
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 255 RASFADRTLAWISRWG--QKKLISAPFLFYPRQFVFEVLLLFQEGGLKALNGQLGATSM 312
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

QY 300 RLSAFQVDTLEIQGVDTLS 318
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 313 NQALEQLRELRISSVQL 331
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
```

```
RESULT 15
LEG7 MOUSE STANDARD; PRT; 135 AA.
ID LEG7 MOUSE
AC 054974;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Galectin-7.
GN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RX MEDLINE=98362649; PubMed=9697310;
RA Magnaldo T., Fowles D., Darmon M.;
RT "Galectin-7, a marker of all types of stratified epithelia.";
RL Differentiation 63:159-168(1998).
CC -1- FUNCTION: COULD BE INVOLVED IN CELL-CELL AND/OR CELL-MATRIX
CC INTERACTIONS NECESSARY FOR NORMAL GROWTH CONTROL (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MAY BE SECRETED BY A NON-CLASSICAL SECRETORY
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF038562; AAB92566.1; -.
DR HSSP; P47929; 2GAL.
DR MGD; MGI:1316742; Lgalst7.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-binding_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
DR Galaptin; Lectin.
KW INIT MET
FT BINDING 69 75 BY SIMILARITY.
FT BINDING 135 AA; 15042 MW; 9C8B3CBBF36C5013 CRC64; BETA-GALACTOSIDE (POTENTIAL).
SQ SEQUENCE

Query Match 16.1%; Score 283.5; DB 1; Length 135;
Best local similarity 43.7%; Pred. No. 1.2e-16;
Matches 55; Conservative 24; Mismatches 46; Indels 1; Gaps 1;

QY 23 IPGLANGMSYVYIQVASSEHMKRFVNFVVGQDPSDVAFHFNPRFDGMDKVNTLQGG 82
DB 9 LPQGVAVGTWIRIGLVDPDQAGRFHVNLLGEEQADALAHFNPRLD-TSEVFNTRKQGG 67
QY 83 KMGSEERRKSNMFKGAFFELFVLAHHYVVVNGNPFVEYGRLLPLQWTHLQVVDGL 142
DB 68 KMGSEERRKGTGTFPGRGQPFVLLATEEGFKAAYGVDDDEYLHFNHRLPRVRLVEVGDV 127
QY 143 QLQGIN 148
DB 128 QLHSLN 133
```

Search completed: January 7, 2003, 16:18:07
Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 16:13:23 ; Search time 38 Seconds

(without alignments)
1132,630 Million cell updates/sec

Title: US-09-988-292-16

Perfect score: 1759
Sequence: 1 MAYVPAPGYPTNPTLPY.....FORVDTLEIGDVTLSYVQI 323

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1759	100.0	323	18	AAW11841	Human galectin-4-1
2	1759	100.0	323	18	AAW05551	Human colon specific
3	1759	100.0	323	19	AAW55503	Human galectin 8.
4	1759	100.0	323	19	AAW46883	Protein sequence e
5	1759	100.0	323	22	AAW24512	C892P predicted am
6	1759	100.0	323	23	AAO15235	Human galectin-4 (
7	1759	100.0	323	23	AAU76211	Human colon specific
8	1759	100.0	329	23	AAO15241	Human galectin-4 (
9	1759	99.8	351	22	AAW73728	Human colon cancer
10	1475.5	83.9	280	19	AAW61367	Human galectin-4.

11	1458	82.9	268	22	AAW75625	Human colon cancer
12	576	32.7	323	21	AAW56802	Human eosinophil c
13	576	32.7	323	23	ABW77853	Amino acid sequenc
14	576	32.7	323	22	ABW61494	Human NF-KB activa
15	575.5	32.7	341	22	ABW97171	Rainbow trout gale
16	574	32.6	355	20	AAW85664	Galectin-9 like pr
17	574	32.6	355	20	AAW06897	Galectin-9 protein
18	574	32.6	355	23	ABW77852	Amino acid sequenc
19	574	32.6	378	22	AAE13847	Human lung tumour-
20	569	32.3	311	19	AAW56504	Human galectin 9.
21	569	32.3	311	23	ABW77854	Amino acid sequenc
22	569	32.3	329	23	AAW97036	Human bladder canc
23	534.5	30.4	358	21	AAW87404	Human PCTA-1 splic
24	534.5	30.4	358	22	AAW85031	Protein encoded by
25	531.5	30.2	316	21	AAW87403	Human PCTA-1 splic
26	531.5	30.2	316	22	AAW85030	Protein encoded by
27	530.5	30.2	317	19	AAW56505	Human galectin 10.
28	530.5	30.2	368	21	AAW87405	Human PCTA-1 splic
29	530.5	30.0	368	22	AAW85032	Protein encoded by
30	527.5	30.0	317	17	AAW03519	Prostate carcinoma
31	513.5	29.2	316	21	AAW87406	Murine PCTA-1, SEQ
32	507.5	28.9	316	16	AAW75702	Rat galectin-8. R
33	454	25.8	300	23	AAW97824	Human cell membran
34	410.5	23.3	264	20	AAW88106	Human galectin-8 p
35	410.5	23.3	264	22	AAW67687	Amino acid sequenc
36	400.5	22.8	94	23	ABW04847	Human ORFX protei
37	383	21.8	278	16	AAW72597	Osteoclast proteot
38	352	20.0	503	22	ABW56581	Drosophila melanog
39	338.5	19.2	314	23	AAW48755	Human adipose tiss
40	338.5	19.2	316	21	AAW70029	Human extracellular
41	327	18.6	168	21	AAW11899	Human colon tumour
42	327	18.6	168	22	AAW24496	Colon tumour relat
43	312	17.7	609	20	AAW82726	Adenovirus PACTSsi
44	312	17.7	609	20	AAW82727	Adenovirus P194N
45	312	17.7	750	20	AAW82728	Adenovirus PMSp194

ALIGNMENTS

RESULT 1	AAW11841	AAW11841 standard; Protein; 323 AA.
ID	AAW11841	
XX	AAW11841:	
AC	20-OCT-1997 (first entry)	
DT	Human galectin-4-like protein.	
XX		
DE		
XX		
KM	galectin-4; lactose binding; stomach; intestine; reagent; drug;	
KW	research; carbohydrate.	
XX		
OS	Homo sapiens.	
XX		
PN	MO9703190-A1.	
XX		
PD	30-JAN-1997.	
XX		
PF	09-JUL-1996; 96WO-DP01899.	
XX		
PR	11-JUL-1995; 95JP-0174778.	
XX		
PA	(SAGA) SAGAMI CHEM RES CENTRE.	
XX		
PI	Kamata K, Kato S, Sekine S, Yamaguchi T;	
XX	WPI: 1997-119046/11.	
DR	N-PSDB: AAT59539.	
XX		
PT	Recombinant human galectin 4-like protein and gene - has	
XX	lactose-binding ability, used in research involving carbohydrate(s)	

PS Claim 1; Page 13-16; 25pp; Japanese.

XX This sequence is a human galectin 4 (a lactose-binding protein)-like
 CC protein. The protein, which is expressed specifically in the human
 CC stomach and intestines, binds lactose and is useful as a drug and as a
 CC reagent for research involving carbohydrates.

XX Sequence 323 AA;

Query Match 100.0%; Score 1759; DB 18; Length 323;

Best Local Similarity 100.0%; Pred. No. 2.6e-158;

Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPARGYQPTNPTLPYYQIPGGNLVGMVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
 DB 1 MAYVPARGYQPTNPTLPYYQIPGGNLVGMVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
 QY 61 AFHFNPRFDGWDKVFNTLOGGKWSSEERKSRMPFKKGAFAFELVFLAHHYKVVVNGNP 120
 DB 61 AFHFNPRFDGWDKVFNTLOGGKWSSEERKSRMPFKKGAFAFELVFLAHHYKVVVNGNP 120
 QY 121 FYEYGHRLPLQMTVHLQVGDGLQLOSNIFIGGQPLRQGPMPMPYPGPGHCHQQLNSLP 180
 DB 121 FYEYGHRLPLQMTVHLQVGDGLQLOSNIFIGGQPLRQGPMPMPYPGPGHCHQQLNSLP 180
 QY 181 TMEGPTTFNPPVYFGRLOGLTARRTIIIGVVPPTGKSFAINFKVSGSGDIALHINPR 240
 DB 181 TMEGPTTFNPPVYFGRLOGLTARRTIIIGVVPPTGKSFAINFKVSGSGDIALHINPR 240
 QY 241 MNGTIVVRNSLLNGSWGSEKKITHNPFPGQFDFLSIRCLDRFKVYANGQHLFDFAHR 300
 DB 241 MNGTIVVRNSLLNGSWGSEKKITHNPFPGQFDFLSIRCLDRFKVYANGQHLFDFAHR 300
 QY 301 LSAFORVDTLEIOGDVTLVSYQI 323
 DB 301 LSAFORVDTLEIOGDVTLVSYQI 323

RESULT 2

AAW06551

ID AAW06551 standard; Protein; 323 AA.

XX AAW06551;

DT 13-MAR-1997 (first entry)

DE Human colon specific gene CSG9 complete polypeptide.

XX Colon specific gene; CSG9; colon cancer; metastasis; diagnosis;

KW therapy; antibody; vaccine.

XX Homo sapiens.

OS

XX

PN WO9639419-A1.

XX

XX

PD 12-DEC-1996.

XX

PF 06-JUN-1995; 95WO-US07289.

XX

PR 06-JUN-1995; 95WO-US07289.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA

PI Rosen CA, Yu G;

XX

XX WPI; 1997-043054/04.

DR

DR N-PSDB; AAT45888.

XX

XX Human colon specific genes and their expression products - detection

PT of which, in non-colon tissue samples, can be used as indication of

PT colon cancer metastasis

XX

XX Claim 8; Fig 9; 60pp; English.

PS

XX Novel polypeptides (AAW06545-53) are encoded by cDNA clones (see also
 CC AAT45880-92) corresponding to 13 human colon specific genes,
 CC designated CSG1, CSG2, etc.; that are primarily expressed in
 CC tissues derived from the colon. Recombinant CSG polypeptides can
 CC be produced in transformed host cells. They are useful diagnostic
 CC markers for colon cancer and for colon cancer metastasis and can
 CC also be used to screen for (ant)agonist cpds. of therapeutic or
 CC diagnostic value. Antibodies raised against the colon-specific
 CC polypeptides may be used to target colon cancer cells or as part
 CC of a colon cancer vaccine.

SQ Sequence 323 AA;

Query Match 100.0%; Score 1759; DB 18; Length 323;

Best Local Similarity 100.0%; Pred. No. 2.6e-158;

Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPARGYQPTNPTLPYYQIPGGNLVGMVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
 DB 1 MAYVPARGYQPTNPTLPYYQIPGGNLVGMVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
 QY 61 AFHFNPRFDGWDKVFNTLOGGKWSSEERKSRMPFKKGAFAFELVFLAHHYKVVVNGNP 120
 DB 61 AFHFNPRFDGWDKVFNTLOGGKWSSEERKSRMPFKKGAFAFELVFLAHHYKVVVNGNP 120
 QY 121 FYEYGHRLPLQMTVHLQVGDGLQLOSNIFIGGQPLRQGPMPMPYPGPGHCHQQLNSLP 180
 DB 121 FYEYGHRLPLQMTVHLQVGDGLQLOSNIFIGGQPLRQGPMPMPYPGPGHCHQQLNSLP 180
 QY 181 TMEGPTTFNPPVYFGRLOGLTARRTIIIGVVPPTGKSFAINFKVSGSGDIALHINPR 240
 DB 181 TMEGPTTFNPPVYFGRLOGLTARRTIIIGVVPPTGKSFAINFKVSGSGDIALHINPR 240
 QY 241 MNGTIVVRNSLLNGSWGSEKKITHNPFPGQFDFLSIRCLDRFKVYANGQHLFDFAHR 300
 DB 241 MNGTIVVRNSLLNGSWGSEKKITHNPFPGQFDFLSIRCLDRFKVYANGQHLFDFAHR 300
 QY 301 LSAFORVDTLEIOGDVTLVSYQI 323
 DB 301 LSAFORVDTLEIOGDVTLVSYQI 323

RESULT 3

AAW56503

ID AAW56503 standard; Protein; 323 AA.

XX AAW56503;

DT 14-SEP-1998 (first entry)

XX Human galectin 8.

DE

XX

XX Galectin 8; lectin; human; autoimmune disease;

KW inflammatory disease; asthma; allergy; melanoma; renal astrocytoma;

KW Hodgkin disease; breast cancer; ovary cancer; prostate cancer;

KW bone cancer; liver cancer; lung cancer; pancreas cancer;

KW spleen cancer; diagnosis; therapy.

XX

OS Homo sapiens.

XX

XX Location/Qualifiers

FT Key 55..101

FT Region /note= "antigenic region"

FT Region 137..162

FT Region /note= "antigenic region"

FT Region 180..193

FT Region /note= "antigenic region"

FT Region 216..266

FT Region /note= "antigenic region"

XX

XX WO9815624-A1.

XX

PD 16-APR-1998.
 XX
 PF 09-OCT-1997; 97WO-US18261.
 XX
 PR 09-OCT-1996; 96MO-US16565.
 PR 09-OCT-1996; 96US-0028093.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Gentz RL, Ni J, Ruben SM,
 DR WPI; 1998-240812/21.
 DR N-PSDB; AAV29785.
 XX
 PT Galectin 8, 9, 10 and 10SV polynucleotides - used for treating
 PT cancer, autoimmune diseases, inflammatory diseases, asthma, and
 PT allergic diseases
 XX
 PS Claim 9; Fig 1, 118pp; English.

CC This polypeptide comprises novel human galectin 8. Its amino acid
 CC sequence was deduced from a human adult small intestine cDNA clone
 CC (see AAV29785). Galectin 8 appears to be mainly expressed in cells
 CC of the colon and small intestine. Novel galectins 8, 9, 10 and 10
 CC splice variant (10SV) are claimed (see AAV56503-06). These can be
 CC obtained using a claimed method by expression in recombinant host
 CC cells. The galectins are used in a claimed method for treating a
 CC cell growth disorder, especially cancer, autoimmune diseases,
 CC inflammatory diseases, asthma, and allergic diseases, in a mammal.
 CC They are especially useful for diagnosis of melanoma, renal
 CC astrocytoma, Hodgkin disease, and breast, ovarian, prostate, bone,
 CC liver, lung, pancreatic and splenic cancers. The invention further
 CC relates to screening methods for identifying agonists and
 CC antagonists of galectin 8, 9, 20 or 10SV activity, and to
 CC diagnostic methods involving estimating levels of galectin 8, 9, 10
 CC or 10SV protein or mRNA.

SO Sequence 323 AA;

Query Match 100.0%; Score 1759; DB 19; Length 323;
 Best Local Similarity 100.0%; Pred. No. 2.6e-158;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPAPGYQPTYNPTLPIYQPIPGSLNVGMSVYIQVASSEHMKRFVNFVVGDDPGSDV 60
 DB 1 MAYVPAPGYQPTYNPTLPIYQPIPGSLNVGMSVYIQVASSEHMKRFVNFVVGDDPGSDV 60
 QY 61 AFHNPREDGMDKVVFNLTLOGKWSSEERKSMPEKGAFAELVFIYLAHYKVVVNGNP 120
 DB 61 AFHNPREDGMDKVVFNLTLOGKWSSEERKSMPEKGAFAELVFIYLAHYKVVVNGNP 120
 QY 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 DB 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 QY 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 DB 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 QY 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIKGYVPTGKSFAINFKVSSGDIALHINPR 240
 DB 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIKGYVPTGKSFAINFKVSSGDIALHINPR 240
 QY 241 MNGGTIVVNSLLNGWSGSEKKITHNPPGQPFDLISRCGLDRPKYVANGQHLDFPAHR 300
 DB 241 MNGGTIVVNSLLNGWSGSEKKITHNPPGQPFDLISRCGLDRPKYVANGQHLDFPAHR 300
 QY 301 LSAFORVDTLEIQGDVTLISYVOI 323
 DB 301 LSAFORVDTLEIQGDVTLISYVOI 323

RESULT 4
 AAM46883
 ID AAM46883 standard; Protein; 323 AA.
 XX
 AC AAM46883;

XX 22-JUN-1998 (first entry)
 DT
 XX
 DE Protein sequence encoded by a colon-specific gene.

XX Colon-specific gene; probe; detection; expression; human;
 KW diagnostic assay; colon cancer; antibody; screening.
 XX

XX Homo sapiens.

XX US5733748-A.

XX 31-MAR-1998.

XX 06-JUN-1995; 95US-0469667.

XX 06-JUN-1995; 95US-0469667.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen C, Yu G;

XX WPI; 1998-229823/20.
 DR N-PSDB; AAV16676.

PT Colon-specific nucleic acids - useful as probes for detecting colon
 PT cancer micrometastases

PS Claim 1; Fig 9; 51pp; English.

CC AAM46876-80 and W4682-85 represent proteins encoded by colon-specific
 CC genes. The polynucleotides encoding these proteins can be used
 CC as probes to detect expression of the corresponding human genes,
 CC e.g. in diagnostic assays for detecting micrometastases of colon cancer.
 CC Recombinant cells containing the polynucleotides can be used to
 CC produce the proteins, in order that antibodies can be raised and
 CC used in further screening or diagnostics.

SO Sequence 323 AA;

Query Match 100.0%; Score 1759; DB 19; Length 323;
 Best Local Similarity 100.0%; Pred. No. 2.6e-158;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPAPGYQPTYNPTLPIYQPIPGSLNVGMSVYIQVASSEHMKRFVNFVVGDDPGSDV 60
 DB 1 MAYVPAPGYQPTYNPTLPIYQPIPGSLNVGMSVYIQVASSEHMKRFVNFVVGDDPGSDV 60
 QY 61 AFHNPREDGMDKVVFNLTLOGKWSSEERKSMPEKGAFAELVFIYLAHYKVVVNGNP 120
 DB 61 AFHNPREDGMDKVVFNLTLOGKWSSEERKSMPEKGAFAELVFIYLAHYKVVVNGNP 120
 QY 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 DB 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 QY 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 DB 121 FVEYGHRLPLQWVTHLQVDDGLQLOSLNIFIGQPLRPGPPMPPYPGPGHCHQOLNSLP 180
 QY 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIKGYVPTGKSFAINFKVSSGDIALHINPR 240
 DB 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIKGYVPTGKSFAINFKVSSGDIALHINPR 240
 QY 241 MNGGTIVVNSLLNGWSGSEKKITHNPPGQPFDLISRCGLDRPKYVANGQHLDFPAHR 300
 DB 241 MNGGTIVVNSLLNGWSGSEKKITHNPPGQPFDLISRCGLDRPKYVANGQHLDFPAHR 300
 QY 301 LSAFORVDTLEIQGDVTLISYVOI 323
 DB 301 LSAFORVDTLEIQGDVTLISYVOI 323

RESULT 5
 AAM24512
 ID AAM24512 standard; Protein; 323 AA.
 XX

AAC24512;
 12-OCT-2001 (first entry)
 C892P predicted amino acid sequence.
 Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 immunogenic; gene therapy; vaccine; colonic cancer.
 Homo sapiens.
 WO200149716-A2.
 12-JUL-2001.
 29-DEC-2000; 2000WO-US35596.
 30-DEC-1999; 99US-0476296.
 10-JAN-2000; 2000US-0480321.
 15-FEB-2000; 2000US-0504629.
 06-MAR-2000; 2000US-0519444.
 19-MAY-2000; 2000US-0575251.
 29-JUN-2000; 2000US-0609448.
 28-AUG-2000; 2000US-0649811.
 (CORI-) CORIXA CORP.
 Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
 King GE, Wang T, Jiang Y;
 WPI; 2001-441847/47.
 Colon tumor associated proteins and nucleic acids useful for the
 prevention, diagnosis and treatment of colonic cancer -
 Claim 2; Page 436-437; 472pp; English.
 The present invention describes colon tumour associated proteins (I) and
 the polynucleotides (II) that encode them. (I) have cytostatic activity,
 (I) and (II) can be used in gene therapy and vaccine production. (I) and
 (II) may be used in the prevention, diagnosis and treatment of diseases
 associated with inappropriate colon tumour associated protein (TCAP)
 expression, such as colonic cancer. For example, (I) and (II) may be
 used to treat disorders associated with decreased expression by
 rectifying mutations or deletions in a patient's genome that affect the
 activity of TCAPs by expressing inactive proteins or to supplement the
 patients own production of them. Additionally, (II) may be used to
 produce the TCAP proteins, by inserting the nucleic acids into a host
 cell culturing the cell to express the protein. (II) and its
 complementary sequences may also be used as DNA probes in diagnostic
 polymease chain reaction (PCR) and hybridisation assays to detect and
 quantitate the presence of similar nucleic acids in samples, and
 therefore which patients may be in need of restorative therapy. (I) may
 also be used as antigens in the production of antibodies against TCAPs
 and in assays to identify modulators of TCAP expression and activity.
 Anti-(I) antibodies and antagonists may also be used to down regulate
 TCAP expression and activity. The anti-(I) antibodies may also be used
 as diagnostic agents for detecting the presence of TCAPs in samples
 (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
 and AA24494 to AA24523 represent nucleotide and amino acid sequences
 given in the exemplification of the present invention.
 Query Match 100.0%; Score 1759; DB 22; Length 323;
 Best Local Similarity 100.0%; Pred. No. 2,6e-158;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MAYVPAFGYQPTNTPLPYQIPGGLNMGMSVYIQGVASEHMKRFFVNFVVGQDPSDV 60
 1 MAYVPAFGYQPTNTPLPYQIPGGLNMGMSVYIQGVASEHMKRFFVNFVVGQDPSDV 60
 61 AFHFNPRFDGKVVNTLQGGKWSSEERKSNPFFKGAAGFELVFLAEHYKVVVNGNP 120

Db 61 AFHFNPRFDGKVVNTLQGGKWSSEERKSNPFFKGAAGFELVFLAEHYKVVVNGNP 120
 QY 121 FYEYGHRLPLQWVTHLQVDGDLQLOSLINIFIGQPLRPOGPPMPPYPGPGHCHQOLNSLP 180
 Db 121 FYEYGHRLPLQWVTHLQVDGDLQLOSLINIFIGQPLRPOGPPMPPYPGPGHCHQOLNSLP 180
 QY 181 TMEGPTTFNPPVPYFGRLOGGLTARTTIIKGYVPTTKSFAINFKVSSGDIALHINPR 240
 Db 181 TMEGPTTFNPPVPYFGRLOGGLTARTTIIKGYVPTTKSFAINFKVSSGDIALHINPR 240
 QY 241 MNGTIVVRNSLLNGSGSEKKITHNPGPGOFFDLIRCLDRKVDKVVANGHLDFAHR 300
 Db 241 MNGTIVVRNSLLNGSGSEKKITHNPGPGOFFDLIRCLDRKVDKVVANGHLDFAHR 300
 QY 301 LSAFQRVDTLEIQGDVTLVYVQI 323
 Db 301 LSAFQRVDTLEIQGDVTLVYVQI 323
 RESULT 6
 AAO15235
 ID AAO15235 standard; Protein; 323 AA.
 XX AAO15235;
 AC AAO15235;
 XX 05-SEP-2002 (first entry)
 DT Human galectin-4 (Clnl14) protein.
 DE Human galectin-4 (Clnl14) protein.
 XX Human; gastrointestinal cancer; stomach cancer; small intestine cancer;
 KW colon cancer; gastrointestinal specific gene; GSG; galectin-4; Clnl14;
 KW carbonic anhydrase I; Clnl15; gastrointestinal cancer marker.
 XX Homo sapiens.
 OS US2002042088-A1.
 PN 11-APR-2002.
 PD 09-MAR-2001; 2001US-0802674.
 PF 09-MAR-2000; 2000US-188061P.
 PR (MACI/) MACINA R A.
 PA (PIDE/) PIDERIT A.
 PA (SUNY/) SUN Y.
 XX Macina RA, Piderit A, Sun Y;
 PI WPI; 2002-507213/54.
 DR N-PSDB; AAL43636.
 DR Diagnosing, monitoring, staging, imaging and treating cancers, e.g.
 PT gastrointestinal cancers such as stomach, small intestine and colon
 PT cancer, associated with the expression of gastrointestinal specific
 PT genes Clnl14 and Clnl15 -
 XX Claim 7; Page 17; 23pp; English.
 The invention comprises a method for diagnosing the presence of
 CC gastrointestinal cancers (e.g. cancers of the stomach, small intestine
 CC and colon) associated with two gastrointestinal specific genes (GSGs).
 CC The two GSGs are human galectin-4 (Clnl14) and human carbonic anhydrase I
 CC (Clnl15). It has been found that Clnl14 and Clnl15 serve as useful
 CC markers in the diagnosis of gastrointestinal cancer. The method of the
 CC invention is useful for detecting, diagnosing, monitoring, staging,
 CC prognosticating, imaging and treating gastrointestinal cancers associated
 CC with the expression of GSGs Clnl14 and Clnl15. The present amino acid
 CC sequence represents the human galectin-4 (Clnl14) protein.
 XX Sequence 323 AA;
 SQ

Query Match 100.0%; Score 1759; DB 23; Length 323;
 Best Local Similarity 100.0%; Pred. No. 2.6e-158;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPAAGYQPTNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFVNFVVGODPGSDV 60
 DB 1 MAYVPAAGYQPTNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFVNFVVGODPGSDV 60

QY 61 AFHFNPRFDGMDKVFNTLQGGKMSSEKRSMPFKGAFFELVFLAHHYKVVVNGNP 120
 DB 61 AFHFNPRFDGMDKVFNTLQGGKMSSEKRSMPFKGAFFELVFLAHHYKVVVNGNP 120

QY 121 FVEYGHRLPLOWVTHLQVDDLOLQSIINFIGQPLRPGGPPMPYPGPGHCHQQLNSLP 180
 DB 121 FVEYGHRLPLOWVTHLQVDDLOLQSIINFIGQPLRPGGPPMPYPGPGHCHQQLNSLP 180

QY 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIIKGYVPTGKSFALNFVSGSDIALHINPR 240
 DB 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIIKGYVPTGKSFALNFVSGSDIALHINPR 240

QY 241 MNGGTIVRNSLLNGSWGSEKKTTHNPFPGQFPDLIRCGLDRFKYVANGQHLFDFAPHR 300
 DB 241 MNGGTIVRNSLLNGSWGSEKKTTHNPFPGQFPDLIRCGLDRFKYVANGQHLFDFAPHR 300

QY 301 LSAFORVDLTLEIOGDVTLSTYQI 323
 DB 301 LSAFORVDLTLEIOGDVTLSTYQI 323

RESULT 7

AAU76211
 ID AAU76211 standard; Protein; 323 AA.

AAU76211;

08-MAY-2002 (first entry)

Human colon specific gene 9 (CSG9) protein.

Human; colon specific gene; CSG; cytostatic; colon disorder; immunogen;

gastric cancer; ulcerative colitis; granulomatous colitis; colon cancer;

cancer cell.

Homo sapiens.

US6337195-B1.

08-JAN-2002.

31-MAR-1998; 98US-0224110.

06-JUN-1995; 95US-0469667.

(HUMA-) HUMAN GENOME SCI INC.

Yu G, Rosen C;

WPI; 2002-163239/21.

Human colon specific gene polypeptide, useful as diagnostic marker,

vaccine and for screening agonists and antagonists for treating colon

cancer -

Example 1; Fig 9; 49pp; English.

This invention relates to the cDNA and protein sequences of novel
 human colon specific genes. These genes which were isolated from a human
 colon tissue cDNA library may have cytosolic activity and may be used
 as a vaccine against colon disorders such as gastric cancer, ulcerative
 colitis and granulomatous colitis. The nucleotide sequences of the
 invention may be used to identify corresponding full length gene
 sequences and is useful as diagnostic marker for colon cancer and as a

CC colon cancer vaccine. The protein sequences of the invention are
 CC useful for preparing monoclonal antibodies specific to CSG proteins,
 CC these antibodies can be used to isolate the protein from colon tissue or
 CC for protein quantitation from a sample using enzyme linked immunosorbant
 CC assay (ELISA). It is also useful for targeting cancer cells and for
 CC screening agonist and antagonist for the CSG proteins which are useful
 CC for treating colon cancer. The present sequence represents the human
 CC colon specific gene sequence (CSG9) protein of the invention.

Sequence 323 AA;

Query Match 100.0%; Score 1759; DB 23; Length 323;
 Best Local Similarity 100.0%; Pred. No. 2.6e-158;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYVPAAGYQPTNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFVNFVVGODPGSDV 60

DB 1 MAYVPAAGYQPTNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFVNFVVGODPGSDV 60

QY 61 AFHFNPRFDGMDKVFNTLQGGKMSSEKRSMPFKGAFFELVFLAHHYKVVVNGNP 120

DB 61 AFHFNPRFDGMDKVFNTLQGGKMSSEKRSMPFKGAFFELVFLAHHYKVVVNGNP 120

QY 121 FVEYGHRLPLOWVTHLQVDDLOLQSIINFIGQPLRPGGPPMPYPGPGHCHQQLNSLP 180

DB 121 FVEYGHRLPLOWVTHLQVDDLOLQSIINFIGQPLRPGGPPMPYPGPGHCHQQLNSLP 180

QY 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIIKGYVPTGKSFALNFVSGSDIALHINPR 240

DB 181 TMEGPPTFNPVPYFGRLOGGLTARRTIIIKGYVPTGKSFALNFVSGSDIALHINPR 240

QY 241 MNGGTIVRNSLLNGSWGSEKKTTHNPFPGQFPDLIRCGLDRFKYVANGQHLFDFAPHR 300

DB 241 MNGGTIVRNSLLNGSWGSEKKTTHNPFPGQFPDLIRCGLDRFKYVANGQHLFDFAPHR 300

QY 301 LSAFORVDLTLEIOGDVTLSTYQI 323

DB 301 LSAFORVDLTLEIOGDVTLSTYQI 323

RESULT 8

AAO15241
 ID AAO15241 standard; Protein; 329 AA.

AAO15241;

05-SEP-2002 (first entry)

Human galectin-4 (Cln114) protein with a histidine purification tag.

Human; gastrointestinal cancer; stomach cancer; small intestine cancer;

colon cancer; gastrointestinal specific gene; GSG; galectin-4; Cln114;

carbonic anhydrase I; Cln115; gastrointestinal cancer marker;

histidine tag.

Homo sapiens.

Synthetic.

Key

Misc-difference 324..329

US2002042088-A1.

11-APR-2002.

09-MAR-2001; 2001US-0802674.

09-MAR-2000; 2000US-186061P.

(MACI/) MACINA R A.

(PIDE/) PIDERIT A.

```

PA (SUNY/) SUN Y.
XX
XX Macina RA, Piderit A, Sun Y;
XX WPI; 2002-507213/54.
XX
XX Diagnosing, monitoring, staging, imaging and treating cancers, e.g.
XX gastrointestinal cancers such as stomach, small intestine and colon
XX cancer, associated with the expression of gastrointestinal specific
XX genes Clnl14 and Clnl15 -
XX
XX Example 2; Page 15; 23pp; English.
XX
XX The invention comprises a method for diagnosing the presence of
XX gastrointestinal cancers (e.g. cancers of the stomach, small intestine
XX and colon) associated with two gastrointestinal specific genes (GSGs).
XX The two GSGs are human galectin-4 (Clnl14) and human carbonic anhydrase I
XX (Clnl15). It has been found that Clnl14 and Clnl15 serve as useful
XX markers in the diagnosis of gastrointestinal cancer. The method of the
XX invention is useful for detecting, diagnosing, monitoring, staging,
XX prognosticating, imaging and treating gastrointestinal cancers associated
XX with the expression of GSGs Clnl14 and Clnl15. The present amino acid
XX sequence represents the human galectin-4 (Clnl14) protein with a
XX C-terminal 6 histidine purification tag.
XX
XX Sequence 329 AA;
XX
XX Query Match 100.0%; Score 1759; DB 23; Length 329;
XX Best Local Similarity 100.0%; Pred. No. 2.7e-158;
XX Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MAYVPAPGYQTYNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFFVNVVGGDPSDV 60
XX DB 1 MAYVPAPGYQTYNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFFVNVVGGDPSDV 60
XX
XX QY 61 AFHNPFRDGDWVKVFNLTQGGKMGSEERKSMFPFKGAAPFVFLAELHYKVVVNGNP 120
XX DB 61 AFHNPFRDGDWVKVFNLTQGGKMGSEERKSMFPFKGAAPFVFLAELHYKVVVNGNP 120
XX
XX QY 121 FVEYGHRLPQWVTHLQVGDGLQSLQINFIGGQPLRPGQPPMPPYPGPGHCHQQLNSLP 180
XX DB 121 FVEYGHRLPQWVTHLQVGDGLQSLQINFIGGQPLRPGQPPMPPYPGPGHCHQQLNSLP 180
XX
XX QY 181 TMEGPPFNPPVPYFGRLOGLTARRTIIKGYVPPTGKSFAINFKVSSGDIALHINPR 240
XX DB 181 TMEGPPFNPPVPYFGRLOGLTARRTIIKGYVPPTGKSFAINFKVSSGDIALHINPR 240
XX
XX QY 241 MNGTIVVNSLLNGSWGSEKKITHNPFPGQPFDLIRCGLDLRFKYYANGQHLFDFFAHR 300
XX DB 241 MNGTIVVNSLLNGSWGSEKKITHNPFPGQPFDLIRCGLDLRFKYYANGQHLFDFFAHR 300
XX
XX QY 301 LSAFORVDTLEIQGDVTLVYQI 323
XX DB 301 LSAFORVDTLEIQGDVTLVYQI 323
XX
XX RESULT 9
XX AAG73728
XX ID AAG73728 standard; Protein; 351 AA.
XX AC AAG73728;
XX
XX DT 03-SEP-2001 (first entry)
XX
XX DE Human colon cancer antigen protein SEQ ID NO:4492.
XX
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX
XX OS Homo sapiens.
XX
XX PN WO200122920-A2.
XX
XX

```

```

PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
XX N-PSDB; AAH33159.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 6292-6294; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAH77789 represent sequences used in the exemplification of the
XX present invention.
XX
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 351 AA;
XX
XX Query Match 99.8%; Score 1755; DB 22; Length 351;
XX Best Local Similarity 99.7%; Pred. No. 7e-158;
XX Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MAYVPAPGYQTYNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFFVNVVGGDPSDV 60
XX DB 29 MAYVPAPGYQTYNPTLPYQPIPGGLNMGMSVYIQGVASEHMKRFFVNVVGGDPSDV 88
XX
XX QY 61 AFHNPFRDGDWVKVFNLTQGGKMGSEERKSMFPFKGAAPFVFLAELHYKVVVNGNP 120
XX DB 89 AFHNPFRDGDWVKVFNLTQGGKMGSEERKSMFPFKGAAPFVFLAELHYKVVVNGNP 148
XX
XX QY 121 FVEYGHRLPQWVTHLQVGDGLQSLQINFIGGQPLRPGQPPMPPYPGPGHCHQQLNSLP 180
XX DB 149 FVEYGHRLPQWVTHLQVGDGLQSLQINFIGGQPLRPGQPPMPPYPGPGHCHQQLNSLP 208
XX
XX QY 181 TMEGPPFNPPVPYFGRLOGLTARRTIIKGYVPPTGKSFAINFKVSSGDIALHINPR 240
XX DB 209 TMEGPPFNPPVPYFGRLOGLTARRTIIKGYVPPTGKSFAINFKVSSGDIALHINPR 268
XX
XX QY 241 MNGTIVVNSLLNGSWGSEKKITHNPFPGQPFDLIRCGLDLRFKYYANGQHLFDFFAHR 300
XX DB 269 MNGTIVVNSLLNGSWGSEKKITHNPFPGQPFDLIRCGLDLRFKYYANGQHLFDFFAHR 328
XX
XX QY 301 LSAFORVDTLEIQGDVTLVYQI 323
XX DB 329 LSAFORVDTLEIQGDVTLVYQI 351
XX
XX RESULT 10
XX AAW61367
XX ID AAW61367 standard; Protein; 280 AA.

```

```
XX AA061367;
AC
XX
XX 05-OCT-1998 (first entry)
DT
XX
XX Human galectin-4.
DE
XX
XX Human; galectin-4; hG4; detection; tumour; breast cancer; diagnosis;
KM cell proliferative disorder; metastasis.
KW
XX Homo sapiens.
OS
XX WO9822139-A1.
PN
XX 28-MAY-1998.
PD
XX 19-NOV-1997; 97WO-US21807.
PF
XX 22-NOV-1996; 96US-0031711.
PR
XX (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.
PA
XX Huflajt ME, Liu F;
PI
XX WPI; 1998-312177/27.
DR N-PSDB; AAV28088.
XX
XX Newly isolated human galectin-4 protein - used to develop products
PT for the diagnosis, study, prevention and treatment of cancers,
PT particularly breast cancer
PS
XX Claim 15; Fig 2; 76pp; English.
XX
XX The present sequence represents human galectin-4 (hG4). Antibodies
CC which bind the hG4 protein, and probes or primers derived from the
CC nucleic acid sequence encoding hG4, can be used for the diagnosis,
CC study, prevention and treatment of hG4-associated disorders,
CC particularly cancers (especially breast cancer). They can also be
CC used for determining the presence of metastases and determining the
CC susceptibility of a subject with a tumour to develop metastases.
CC
XX
SQ Sequence 280 AA;
Query Match 83.9%; Score 1475.5; DB 19; Length 280;
Best Local Similarity 97.5%; Pred. No. 1.6e-131;
Matches 273; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
QY 44 KRFFVNFVVGDPGSDVAHFNFDPGMDKVVFNILQGKMGSEERKSMPPKGAAPFL 103
DB 2 KRFFVNFVVGDPGSDVAHFNFDPGMDKVVFNILQGKMGSEERKSMPPKGAAPFL 61
QY 104 VFIVLAEHKVVVNGNPFYEGHRLPLQWVTHLQVNDGLQLOSINFIGQPLRPGCPMM 163
DB 62 VFIVLAEHKVVVNGNPFYEGHRLPLQWVTHLQVNDGLQLOSINFIGQPLRPGCPMM 121
QY 164 PYPGSGHGHQOLNLSPTMEGPPPTFNPVPYFGRLQGLTARTIIIKGYVPTGKSPAI 223
DB 122 PYPGSGHGHQOLNLSPTMEGPPPTFNPVPYFGRLQGLTARTIIIKGYVPTGKSPAI 181
QY 224 NFKVSSGGDIALHINPRMGNGTVVRNSLNGSGSEKKITNHPFGQGFPLSIRCGID 283
DB 182 NFKVSSGGDIALHINPRMGNGTVVRNSLNGSGSEKKITNHPFGQGFPLSIRCGID 240
QY 284 RKFVYANGQHLFDPFAHRLSAFQRYVDLLEIQGDVTLSTYQI 323
DB 241 RKFVYANGQHLFDPFAHRLSAFQRYVDLLEIQGDVTLSTYQI 280
RESULT 11
AA075625
ID AA075625 standard; Protein; 268 AA.
XX
XX AA075625;
```

```
XX 03-SEP-2001 (first entry)
DT
XX
XX Human colon cancer antigen protein SEQ ID NO:6389.
DE
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma.
KW
XX Homo sapiens.
OS
XX WO200122920-A2.
PN
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI; 2001-235357/24.
DR N-PSDB; AAH35030.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT
XX Claim 11; Page 7869-7871; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AA073514 to AA077788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH7789 represent sequences used in the exemplification of the
CC present invention.
CC
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 268 AA;
Query Match 82.9%; Score 1458; DB 22; Length 268;
Best Local Similarity 100.0%; Pred. No. 6.9e-130;
Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 PGSDVAHFNFDPGMDKVVFNILQGKMGSEERKSMPPKGAAPFLVIVLAEHKVV 115
DB 1 PGSDVAHFNFDPGMDKVVFNILQGKMGSEERKSMPPKGAAPFLVIVLAEHKVV 60
QY 116 VNGNPFYEGHRLPLQWVTHLQVNDGLQLOSINFIGQPLRPGCPMMPPYPGSGHGHQO 175
DB 61 VNGNPFYEGHRLPLQWVTHLQVNDGLQLOSINFIGQPLRPGCPMMPPYPGSGHGHQO 120
QY 176 LNSLPTMEGPPPTFNPVPYFGRLQGLTARTIIIKGYVPTGKSPAINFKVSSGGDIAL 235
DB 121 LNSLPTMEGPPPTFNPVPYFGRLQGLTARTIIIKGYVPTGKSPAINFKVSSGGDIAL 180
QY 236 HINPRMGNGTVVRNSLNGSGSEKKITNHPFGQGFPLSIRCGIDRKYVANGQHLF 255
DB 181 HINPRMGNGTVVRNSLNGSGSEKKITNHPFGQGFPLSIRCGIDRKYVANGQHLF 240
```

QY 296 DEAHLSAFORVDTLEIQGVDTLSYVQI 323
 Db 241 DEAHLSAFORVDTLEIQGVDTLSYVQI 268

RESULT 12

AAV56802
 ID AAV56802 standard; Protein; 323 AA.

XX AAV56802;

DT 27-MAR-2000 (first entry)

DE Human eosinophil chemotactic factor (ealelectin).

XX Eosinophil chemotactic factor; galectin; chemotaxis inhibitor;
 KW inflammatory disorder; eosinophil; allergic disease; asthma; human;
 KW allergic rhinitis; atopic dermatitis; ealelectin; galectin.

XX Homo sapiens.

XX WO9962556-A1.

XX 09-DEC-1999.

XX 02-JUN-1999; 99WO-JP02952.

XX 03-JUN-1998; 98JP-0170698.

XX (EFFE-) EFFECTOR CELL INST.

PI Kanegasaki S, Matsumoto R, Hirashima M;

DR WPI; 2000-086860/07.

DR N-PSDB; AAZ46748.

XX Ecalectin and other galectins as eosinophil chemotaxis promoters for
 PT screening potential inhibitors of this effect -

PS Disclosure; Page 35-39; 49pp; Japanese.

XX The invention provides compositions for increasing the chemotacticity of
 CC eosinophils that contain as active component a galectin. Chemotaxis
 CC inhibitors identified by the screening are used as agents for the
 CC treatment and prevention of inflammatory disorders caused by the increase
 CC of eosinophil content in affected tissues, such as allergic diseases
 CC including bronchial asthma, allergic rhinitis and atopic dermatitis. The
 CC present sequence represents a human eosinophil chemotactic factor,
 CC ecalectin. Ecalectin belongs to the family of galectin and has
 CC eosinophil chemotactic activity.

XX Sequence 323 AA;

Query Match 32.7%; Score 576; DB 21; Length 323;

Best Local Similarity 40.1%; Pred. No. 3.3e-46;

Matches 131; Conservative 50; Mismatches 118; Indels 28; Gaps 9;

QY 11 PTYNPTLPYQPIPGGLNVGMVYIQG-VASEHMKRFFVNVVVGQDPSDVAFHFNPRFD 69
 Db 9 PYLSPAVPFSGTIQGLDGLQITVNGTVLSSGTRFAVNFQTGFS-GNDIAPHFNPRFE 67

QY 70 GWDKVVNTLOGKNGSEERKSNPFKGAAPFELFVLAEHYKVNVNGNPFVEYGHRLP 129
 Db 68 DGGYVVCNTRQNGSWGPEERKTHMPFOKMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP 127

QY 130 LQWYTHLQVDDQLQIQINIFIGGQLRPGQPPMMPPVPGP-----GHCHQ-----QLNSLP 180
 Db 128 FHRVDTISVNGSVQLSYSP-----QPPGVWPNAPAITQTVIHTVQSAPGQMFSTP 179

QY 181 TMGEPPTFNP-----PVPYFGRLOGLPARRTIKGVVPTGKSFAINFKVSGSGDIALH 236
 Db 180 AI--PPMYPHAPYMPFITILGGLYPSKILLSGTVLPSAQRFIN--LCSGNHIAFH 235

QY 237 INPRMGNGTVVRNLLNGSWGSEBKITHN-PFGPGOFFDLIRCGLDLRFKVVANGOHLF 295
 Db 236 LNPRFDENAVVRNTQIDNSWGSEERSLPKMPFVRGQSFVWILCEAHCLKVAVDGQHLF 295
 QY 296 DPAHLSAFORVDTLEIQGVDTLSYVQ 322
 Db 296 EYHRLRLNPLPTINRLEVGGDIQLTHVQ 322

RESULT 13

ABB77853

ID ABB77853 standard; protein; 323 AA.

XX ABB77853;

DT 27-SEP-2002 (first entry)

DE Amino acid sequence of a human protein.

XX Cancer; galectin 9; antibody; skin cancer; melanoma; breast cancer;
 KW ovarian cancer; uterus cancer; tumour; prostate cancer; bladder cancer;
 KW kidney cancer; thyroid cancer; throat cancer; tongue cancer;
 KW upper jaw cancer; esophageal cancer; stomach cancer; colon cancer;
 KW lung cancer; liver cancer; gall-bladder cancer; pancreatic cancer;
 KW leukemia; liposarcoma; glioma.

XX Homo sapiens.

XX WO200237114-A1.

XX 10-MAY-2002.

XX 31-OCT-2001; 2001WO-JP09561.

XX 01-NOV-2000; 2000JP-0335077.

XX (GALP-) GALPHARMA CO LTD.

XX Hirashima M, Yamauchi A, Kageshita T, Nakamura T, Nishi N;

XX WPI; 2002-519265/55.

XX Metastasis mechanism-based agents (anti-galectin 9 antibody) for
 PT detecting ability of cancer to metastasize in cells, uses galectin 9 as
 PT marker to detect cancer metastasis for diagnosis, cancer prevention and
 PT treatment -

XX Example 5; Page 61-63; 68pp; Japanese.

XX The specification describes an agent for detecting the ability of
 CC cancers to metastasize. This agent comprises anti-galectin 9
 CC antibody as the active ingredient. The agents and method are for or
 CC detecting ability of cancer to metastasize in cancer cells, which are
 CC useful in the diagnosis, prevention and treatment of cancer. The cancer
 CC includes epithelial malignant tumours (tumourising or non-tumourising)
 CC in organs, tissues or blood. Such cancer can be 1 of the not less than 31
 CC specifically claimed, e.g. skin cancer (including melanoma), breast
 CC cancer, ovarian cancer, uterus cancer, malignant tumour of the testis,
 CC prostate cancer, bladder cancer, kidney cancer, thyroid cancer, cancer
 CC of throat and larynx, tongue cancer, upper jaw cancer, esophageal
 CC cancer, stomach cancer, colon and rectum cancer, lung and bronchus
 CC cancer, liver cancer (including hepatocarcinoma and intrahepatic
 CC biliary cancer), extrahepatic biliary duct and gall-bladder cancer,
 CC pancreatic cancer, leukemia, malignant lymphoma, liposarcoma and
 CC glioma. The cancer cells are particularly breast cancer cells or
 CC melanoma cells. The present sequence represents a human protein,
 CC which is used in the course of the invention.

XX Sequence 323 AA;

Query Match 32.7%; Score 576; DB 23; Length 323;

Best Local Similarity 40.1%; Pred. No. 3.3e-46;

Matches 131; Conservative 50; Mismatches 118; Indels 28; Gaps 9;

